Fisheries and Oceans Canada | Pêches et Océans Canada
Science | Sciences

Canadian Science Advisory Secretariat (CSAS)

Research Document 2013/046

Central and Arctic Region

Genetic differentiation and origin of the Shortjaw Cisco (*Coregonus zenithicus*) in the Great Lakes and other inland Canadian lakes

Julie Turgeon and Audrey Bourret

Département de biologie
Université Laval
Quebec City, Quebec
Canada G1V 0A6

November 2013

Canada

Foreword

This series documents the scientific basis for the evaluation of aquatic resources and ecosystems in Canada. As such, it addresses the issues of the day in the time frames required and the documents it contains are not intended as definitive statements on the subjects addressed but rather as progress reports on ongoing investigations.

Research documents are produced in the official language in which they are provided to the Secretariat.

Published by:

Fisheries and Oceans Canada
Canadian Science Advisory Secretariat
200 Kent Street
Ottawa ON K1A 0E6

http://www.dfo-mpo.gc.ca/csas-sccs/
csas-sccs@dfo-mpo.gc.ca

© Her Majesty the Queen in Right of Canada, 2013
ISSN 1919-5044

Correct citation for this publication:

Turgeon, J., and Bourret, A. 2013. Genetic differentiation and origin of the Shortjaw Cisco (*Coregonus zenithicus*) in the Great Lakes and other inland Canadian lakes. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/046. iv + 37 p.

TABLE OF CONTENTS

ABSTRACT IV

RÉSUMÉ IV

INTRODUCTION 1

MATERIALS AND METHODS 1

- BIOLOGICAL SAMPLES 1
- GENETIC CHARACTERIZATION WITH AFLP MARKERS 2
- DATA ANALYSIS 2

RESULTS 4

- LARGE SCALE PATTERN (TOTAL DATASET) 4
- GREAT LAKES 4
- SHORTJAW CISCO DATASET 5
- INLAND LAKES – SHORTJAW CISCO VS. SYMPATRIC MORPHOTYPES 5
  - White Partridge Lake 5
  - Trout Lake 6
  - Brule Lake 6
  - Lake of the Woods 6
  - Lake Athapapuskow 6
  - Great Slave Lake 7
  - Great Bear Lake 7

DISCUSSION 7

ACKNOWLEDGEMENTS 9

REFERENCES CITED 9

ABSTRACT

Ciscoes display a phenomenal level of ecophenotypic diversity throughout their North American range, leading to taxonomic uncertainty and complicating conservation efforts. Predictions associated with three hypotheses on the origin of this diversity, and in particular of the Shortjaw distinct phenotype, are evaluated. These hypotheses are the 'Plasticity Hypothesis', the 'Good Species Hypothesis', and the 'Parallel Origin Hypothesis'. Patterns of genetic variation at 290 AFLP loci among 1371 individuals from twenty lakes are analysed, including 387 individual fish identified as (or likely representing) Shortjaw Cisco (*Coregonus zenithicus*) from 10 lakes. Genetic cluster analyses, association between individual genetic characteristics and phenotypic attributes, genetic re-allocation and analyses of molecular variance were performed. Evidence for the genetic distinctiveness of the Shortjaw Cisco was strong in Lake Nipigon, Trout Lake, Athapapuskow Lake and Great Bear Lake, weak in White Partridge Lake and Lake Superior, and absent in Brule Lake, Lake of the Woods, and Great Slave Lake. It could not be tested in Lake Huron. The Plasticity Hypothesis is dismissed given genetic distinctiveness of morphotypes within many lakes. There is no evidence that Shortjaw Cisco form a distinct lineage, contrary to the predictions of the Good Species Hypothesis. Shortjaw Cisco were always more closely related to sympatric forms than to Shortjaw Cisco from allopatric locations, in accordance with the Parallel Origin Hypothesis. From a genetic and evolutionary standpoint, the Shortjaw Cisco has multiple origins and is diagnosable only at the local scale.

Différences génétiques et origine du cisco à mâchoires égales (*Coregonus zenithicus*) vivant dans les Grands Lacs et d'autres lacs intérieurs canadiens

RÉSUMÉ

Les ciscos présentent dans leur aire de répartition en Amérique du Nord un niveau phénoménal de diversité écophénotypique, ce qui donne lieu à de l'incertitude taxonomique et complique les activités de conservation. On a évalué les prédictions liées à trois hypothèses quant à l'origine de cette diversité, notamment en ce qui a trait au phénotype distinct du cisco à mâchoires égales. Ces hypothèses sont l'« hypothèse de la plasticité », l'« hypothèse de l'espèce valide » et l'« hypothèse de l'origine parallèle ». On a analysé les tendances de la variation génétique au loci d'AFLP 290 chez 1 371 individus provenant de vingt lacs. Les spécimens analysés comprenaient 387 individus identifiés comme étant des ciscos à mâchoires égales (*Coregonus zenithicus*) (ou comme s'y apparentant) provenant de dix lacs. On a analysé les groupes génétiques, la corrélation entre les caractéristiques génétiques individuelles et les attributs phénotypiques, la redistribution génétique et la variance moléculaire. Les preuves d'une différenciation génétique chez le cisco à mâchoires égales étaient importantes dans les lacs Nipigon, Trout, Athapapuskow et Great Bear, faibles dans les lacs White Partridge et Supérieur, et absentes dans le lac Brule, le lac des Bois et le Grand lac des Esclaves. La différenciation génétique n'a pu être testée dans le lac Huron. On rejette l'hypothèse de la plasticité, étant donné la différenciation génétique des morphotypes dans bon nombre de lacs. Il n'y a aucune preuve démontrant que le cisco à mâchoires égales constitue une lignée distincte, ce que suggère l'hypothèse de l'espèce valide. Les ciscos à mâchoires égales ressemblaient toujours davantage aux formes sympatriques qu'aux individus provenant de sites allopatriques, conformément à l'hypothèse de l'origine parallèle. D'un point de vue génétique et évolutif, le cisco à mâchoires égales présente des origines multiples et n'est identifiable qu'à l'échelle locale.

INTRODUCTION

Ciscoes display a phenomenal level of ecophenotypic diversity throughout their North American range (Scott and Crossman 1973; Clarke 1973). This diversity has been and still is impeding taxonomic designations of individual fish found in lakes where more than one type of cisco occurs. Given the rarity of certain types of ciscoes, as well as their likely distinct role in lacustrine ecosystems, this taxonomic uncertainty is complicating conservation efforts. At the same time, the origin of this diversity remains contentious. The simplest hypothesis is that there is only one lineage of ciscoes. Within lakes with two or more cisco types, the phenotypic diversity would then merely be the result of plasticity of fish experiencing different lake environments (the **Plasticity Hypothesis**). Another hypothesis is that each (or most) species currently recognized by experts (e.g., Scott and Crossman 1973) or previously described (e.g., Koelz 1929) represents a genetically and demographically independent lineage, as most good species do (the **Good Species Hypothesis**). Finally, the **Parallel Origin Hypothesis** proposes that the various types of ciscoes originated very recently (i.e., since the last glacier retreat, ca. 10 Kya), independently and repeatedly in (or near) each of the lakes where they are currently found (Clarke 1973; Smith and Todd 1984; Turgeon and Bernatchez 2003). According to Turgeon and Bernatchez (2001a,b, 2003), patterns of genetic variation at seven microsatellite loci among ciscoes from 22 lakes support the postglacial parallel in-lake divergence of several morphotypes following the admixture of two glacial races. The parallel evolution of eco/morphotypes is not uncommon in the fish fauna of the North American (and Eurasia) (e.g., Schluter 1996; Hendry 2009), with strong evidence supporting this scenario for whitefish (Bernatchez 2004) and sticklebacks (Taylor and McPhail 2000; Rundle et al. 2000).

This study examines whether patterns of genetic variation among cisco morphotypes fit the predictions associated with the above hypotheses (see Materials and Methods for predictions). Special attention is given to evidence supporting the genetic distinctiveness of ciscoes that are officially or putatively identified as Shortjaw Cisco (SJ). The Laurentian Great Lakes (GL) are analyzed separately from the inland lakes because they once harboured a very rich cisco fauna (Koelz 1929). Ecological perturbations have led to the extinction of several forms, and possibly to the hybridization of others (Todd and Stedman 1989). Nevertheless, lakes Superior and Nipigon currently each harbour four types of ciscoes (Turgeon et al. 1999; Pratt 2012, Pratt and Chong 2012; Pratt 2013). In contrast, the cisco fauna of relatively pristine inland lakes is less diverse (Scott and Crossman 1973; Clarke 1973). Most lakes only harbour the Lake Cisco (LC), a fair number of lakes are known to comprise both the Lake Cisco and the Shortjaw Cisco, and multiple forms are reported in Great Slave Lake (Muir et al. 2013).

MATERIALS AND METHODS

BIOLOGICAL SAMPLES

Tissues were obtained from 20 lakes throughout the range of ciscoes in Canada (Table 1, Figure 1). We used cisco common names or the description based on phenotypic attributes provided with samples to define morphotypes (MT) (Table 1). Samples were obtained for all extant types of ciscoes reported from the Great Lakes (except Lake Cisco from Lake Erie). Several individuals representing Shortjaw Cisco were obtained from lakes Nipigon and Superior, but only two individuals were available from Lake Huron. SJ were obtained from a large proportion of the few inland lakes where it has been reported (i.e., White Partridge Lake, Trout Lake, Brule Lake, Lake of the Woods, Athapapuskow Lake, Great Slave Lake, and Great Bear Lake). In Great Bear Lake, ciscoes captured in deep waters were classified as Shortjaw Cisco as they show morphological characteristics consistent with this description (Howland et al.

2013). Samples from lakes where only one type of cisco was reported to occur, in general LC, were also analysed to assess the large scale genetic structure.

For several lakes, information on individual phenotypes was provided with samples. In most cases, this concerned the total number of gill rakers (GR) and/or total body length. For Lake Nipigon, individual scores from multivariate analyses based on morphometric traits were used (S. Reid, pers. comm.). For Great Bear Lake, the depth and lake sector where fish were sampled were also considered given that the SJ-like morphotype displayed phenotypic variation between lake sectors (Howland et al. 2013).

GENETIC CHARACTERIZATION WITH AFLP MARKERS

All tissues samples (muscles or fins) were preserved in 95 % EtOH. DNA was extracted from approximately 25 mm^2 tissues with the QIAGENE blood and tissues extraction kit (QIAGEN), and DNA quality was assessed on 2% agarose gels. Only good quality DNA was retained (high concentration, no sign of degradation). AFLP fragments were generated following AFLP® Plant Mapping protocol of Applied Biosystems (2007-2010). Approximately 200 ng of DNA were used for the restriction-ligation step with *Eco*RI and *Mse*I (New England Biolabs, Ipswich, MA). Six *Eco*RI/*Mse*I primer pairs were used for the selective PCRs (Table 2). Selective PCRs were slightly modified in that i) primers comprised 4 nucleotides on top of the *Mse*I sequence and ii) cycling included a denaturation step of 20 s at 94°C, 9 'step-down' cycles with 30 s annealing step beginning at 69°C and ending at 61°C, 20 cycles with 30 s annealing step at 60°C, and a final 2 min extension step at 72°C. PCR products were co-migrated on an ABI 3100 or ABI 3130XL capillary sequencer alongside the LIZ 500 size standard (Applied Biosystems). Positive controls were always included to ensure inter-run compatibility and data quality. AFLP profiles were checked manually and scored using GENEMAPPER 3.7 analysis software (Applied Biosystems). Only those peaks with a minimum relative fluorescence of 100 units were considered.

We amplified a total of 372 AFLP loci, of which 290 were polymorphic using a 5% criterion within sample (i.e., one MT in one (sector of a) lake). We replicated 92 genotypes (6.7%) from the restriction step, yielding a genotyping error rate of 2.1% for polymorphic loci (Bonin et al. 2004).

As indicated in Table 1, the total dataset includes genotypes for 1371 individuals. The Great Lakes dataset includes 647 individuals, of which 126 are SJ. The Shortjaw dataset includes samples of SJ and LC from 8 lakes where they co-occur, with Lake Huron being excluded given the insufficient number of SJ samples (N=2, Table 1). In all, this leaves 371 LC and 320 SJ (including the 60 fish from White Partridge Lake that are putative SJ).

DATA ANALYSIS

The three hypotheses presented above yield contrasting predictions on patterns of genetic resemblance and dissimilarity between MTs within and among lakes. The **Plasticity Hypothesis** predicts geographical structure within one general cisco lineage (or perhaps two lineages representing glacial races, Turgeon and Bernatchez 2001a). Genetic clusters should then correspond to (groups of nearby) lakes. Most importantly, this hypothesis predicts that MTs do not form distinct gene pools within lakes. The **Good Species Hypothesis** predicts that there should be several independent lineages, each representing a morphospecies. One such lineage must correspond to SJ. Genetic clusters should correspond to MTs, and geographical structure should be nested within MTs. Samples of a given MT from different lakes should always be more similar to one another than to samples of another MT occurring in the same (or a different) lake. In summary, taxonomy (i.e., MTs) should govern patterns of genetic variation. The **Parallel**

Origin Hypothesis predicts that several regional lineages should comprise pairs (or groups) of distinct MTs. Genetic clusters should correspond to locale (lakes or regions), and morphological variation should be nested within locale. Samples of different MTs from the same lake should be more similar to one another than to samples of the same MT occurring in a different lake. Depending on how far *in situ* differentiation has proceeded, genetic differentiation between sympatric MTs may range from undetectable to highly significant (Hendry 2009). In summary, geography should govern patterns of genetic variation, with morphotypes such as SJ being genetically supported only on small scales. Note that more complex evolutionary histories can be envisioned (Hudson et al. 2007), but most are not amenable to clear predictions at this point in time. We prefer to restrict our analyses to those three main hypotheses because we can make distinct sets of predictions.

Our first analyses were performed with algorithms that form K genetic clusters while using no *a priori* information on the morphological (MTs) or geographical (lake of origin) attributes of individual genotypes. We used two algorithms relying on different clustering principles, i.e., STRUCTURE v.2.3.3 (Pritchard et al. 2000; Falush et al. 2007; Hubisz et al. 2009) and FLOCK v.2.0 (Duchesne and Turgeon 2009, 2012). STRUCTURE is a Bayesian method forming clusters that best respect Hardy-Weinberg genotypic proportions and linkage equilibrium. Once the clusters have been formed, this method provides a coefficient of ancestry for each individual (*q*-values). There is no absolute rule to assign individuals to clusters on the basis of *q*-values, but values smaller than 0.3 or larger than 0.7 are customary. We set burn-in to 50,000 iterations and subsequent run lengths to 200,000 iterations. We did 10 runs for each *K* value tested. We used Ln P(X|K) (Pritchard et al. 2000) and ΔK (Evanno et al. 2005) as criteria to infer the most likely number of clusters (*K*). At times, more than one clustering solution was very likely; comparisons with FLOCK were then especially useful. Figures were made with DISTRUCT (Rosenberg 2004). FLOCK is a frequentist and partly deterministic method relying on iterative re-allocation. Starting from a random division of the total sample into *K* partitions, FLOCK iterates re-allocations until cluster compositions are stable. Individuals are allocated to the cluster where their genotype is more likely to occur. Plateau analyses based on the repetition of identical cluster solutions are used to determine *K* (Duchesne and Turgeon 2012). FLOCK run time is much shorter than Structure, so we used this approach hierarchically to find clusters within clusters, as well as a point of comparison with Structure when the most likely number of clusters (K) was difficult to determine. With either algorithm, once the clusters are formed, the phenotypic attributes and the geographical provenance (lake, lake sectors) of individuals belonging to each cluster can be examined. Here, phenotypic attributes were examined in relation to ancestry estimates (q-values) when there was more than one cluster supported in lakes where the SJ was occurring.

A second set of analyses was performed that used the *a priori* classification of fish, namely, genetic re-allocation and analyses of molecular variance (AMOVA). For the re-allocation procedure, an individual fish was assigned to the reference group where its genotype is most likely to occur using AFLPOP (Duchesne and Bernatchez 2002). Reference groups corresponded to MTs or to individual lakes. Eastern and western clusters of lakes identified by STRUCTURE and FLOCK (see Results) were also used as reference groups with the Shortjaw dataset. Re-allocation is a nice complement to cluster analyses: clusters are formed on the basis of group properties, while allocation depends on individual genotypes. Nested AMOVAs were performed with ARLEQUIN 3.5 (Excoffier and Lischer 2010) for the Great Lakes, and for lakes where different sectors were sampled (Athapapuskow Lake, Great Slave Lake, Great Bear Lake). The objective was to determine whether phenotypic attributes (MT) or geographic provenance (Lake or Lake sector) was explaining a larger proportion of the observed genetic variance. We tested a model using *MT* as the main factor, with *Lake* (or *Sector*) nested within *MT*, as well as a model using the reversed nesting scheme, i.e., *Lake* (or *Sector*) as the main factor, and *MT*

nested within *Lake*. The explanatory powers of the two models were compared using the corrected Akaike Information Criterion (AICc) following Halverson et al. (2008).

RESULTS

LARGE SCALE PATTERN (TOTAL DATASET)

Using the entire dataset, the analysis with STRUCTURE provides evidence that samples are primarily divided in two clusters (Table 3, Figure 2A). The eastern cluster comprises all lakes east of Lake Nipigon, including the upper Great Lakes but excluding Scorch Lake. The western cluster includes Lake Nipigon and all lakes west, as well as Scorch Lake. Scorch Lake is located north of Lake Huron near Timmins (northern Ontario) in the Moosonee River drainage flowing into James Bay (Figure 1). The next most likely cluster solution found by STRUCTURE is that of 15 clusters (Figure 2B). Only one cluster corresponds to a MT, namely the Shortjaw Cisco in Lake Nipigon. Twelve other clusters each comprise a single or a group of nearby lakes. These are, roughly from east to west: 1-White Partridge Lake, 2-Trout Lake, 3- Lake Simcoe and Lake Ontario, 4- Manitou, Grand, and Biggar lakes, 5-Kennebec Lake, 6-Brule Lake, 7-Scorch Lake, 8-Upper GL (Huron/Michigan/Superior), 9-Lake Nipigon (SJ excluded), 10- Lake of the Woods, 11-L. Athapapuskow/Churchill River, and 12- Great Slave/Great Bear lakes. The last two clusters cannot be associated with any group of fish. Using FLOCK, two principal clusters are also apparent, following the exact same east vs. west division (Table 3, Figure 3). Subdividing clusters in a hierarchical manner leads to groups that are very similar to the STRUCTURE clusters (Figure 3). The hierarchy of clusters nearly always involved separating lakes, with more than 90% of fish from a given lake (no matter what MT) belonging to the same cluster. As with Structure, a cluster corresponding to Shortjaw was detected only in Lake Nipigon. In Lake of the Woods, fish were parted in two clusters, one of which included the few SJ from this lake (N=8), along with many LC.

GREAT LAKES

Using the Great Lakes dataset, there is evidence for differentiation among lakes, and evidence for genetic differentiation of SJ from other MTs in Lake Nipigon (Table 3, Figure 2C,D, Figure 3). Using STRUCTURE with Evanno's criterion leads to recognizing two clusters, i.e., Lake Nipigon vs. all other lakes (Figure 2C). This is coherent with the overall pattern whereby Lake Nipigon belongs to the 'western' group' while all other GLs are in the eastern group (Figure 2A). Applying Pritchard's criterion yields eight clusters (Figure 2D), six of which can be associated with groups of fish. A first cluster clearly comprises fish from Lake Ontario. A second cluster corresponds to SJ from both sectors of Lake Nipigon. A third cluster more loosely comprises Blackfin Cisco (*C. nigripinnis*) from both sectors of Lake Nipigon. A fourth cluster is associated with the other MTs in Lake Nipigon (Bloater and Lake Cisco). Finally, two clusters are found across the Upper Lakes, with no obvious associations with MTs. FLOCK identifies the same two principal clusters, with further hierarchical analyses corresponding to Figure 3. Fish from Lake Ontario form a cluster apart from that including the upper Great Lakes, and SJ form a cluster within Lake Nipigon.

Re-allocation of individual fish genotypes to each of the Great Lakes is much more successful than re-allocation to MTs (Table 4A vs. 4B). Re-allocation to Lake Nipigon and Lake Ontario is nearly perfect (>95%), and very strong to Lake Superior (75%). However, fish from Lake Huron, and especially those of Lake Michigan, are re-allocated across all three upper lakes. Re-allocation to MTs is successful only for MTs unique to a single lake (Kiyi, *C. kiyi* in Lake Superior and Blackfin in Lake Nipigon, 78% each, Table 4B). More than half of the SJ are re-allocated to other MTs. Bloaters and Lake Cisco are poorly re-allocated (<50%) and often

intermixed. AMOVAs corroborate that genetic variation is primarily associated with differences among lakes (Table 5A, ΔAIC = 29.6). When used as a main factor, *Lake* does explain a significant portion of the variation (9.27 %, P <0.001), while *MT* does not (1.5%, P = 0.107).

Within Lake Nipigon, there is clear evidence for the genetic distinctiveness of the Shortjaw Cisco. STRUCTURE and FLOCK find two clusters (Figure 2E), with SJ from both sectors clearly dominating one cluster. Structure finds another solution at K= 6 (Fig. 2F). It shows that one cluster comprises most SJ, and also suggests that Blackfin Cisco from both lake sectors form another cluster. The remaining clusters are not associated with either Bloaters or Lake Cisco. Individual scores on DF1 for Shortjaw Cisco and Lake Cisco from both lake sectors are plotted against STRUCTURE *q*-values in Figure 4A. SJ tend to have higher *q*-values and higher DF1 score values, but those scores span a wide range. Lake Cisco individuals have lower but highly variable *q*-values, associated with a smaller range of lower DF1 values. The re-allocation matrix (Table 4C) supports the genetic distinctiveness of SJ, with 85% of correct re-allocation. While non-negligible, the proportions of correct re-allocations are much lower for the other MTs. In AMOVAs, *MT* explains a significant portion of the genetic variation when used as a main factor (3.4%, P = 0.006), contrary to *Sector* (0.17%, P = 0.416) (Table 4B, ΔAIC = 8.1).

In Lake Superior, there is very little support for the presence of distinct MTs. Neither STRUCTURE nor FLOCK find evidence for two clusters. The application of Evanno's criterion leads to the minimum possible value of K=2 but there are no clear correspondence between *q*-values and morphotypes or lake sectors (Figure 2G). The re-allocation matrices for each lake sector shows that MTs are poorly re-allocated, except perhaps for SJ in one (northwest) lake sector (Table 4D). In AMOVAs, the proportion of variation explained when *MT* or *Sector* are used as main factors are low and only approaching significance (*MT*: 1.33%, P = 0.05; *Sector*: 1.02%, P = 0.08, Table 5C). Model selection indicates that genetic structure is better explained with a model nesting sectors within morphotypes (ΔAIC = 4.3).

There is no evidence for more than one cluster in Lake Huron. The solution for K = 2 identified with Evanno's criteria in STRUCTURE clearly does not have any biological meaning (not shown).

SHORTJAW CISCO DATASET

There is no evidence that the SJ represents a distinct lineage. SJ from the eastern cluster are genetically more similar to LC from this same cluster than to SJ from the western cluster (Table 6). Indeed, fish are very poorly allocated to MTs across the east/west spatial boundary, but they are perfectly allocated into the same spatial group across the MT boundary. Nearly all fish from the eastern group, whether SJ or LC, are allocated to western LC (Table 6A, east to west), while SJ and LC from western sites are allocated nearly equally between eastern SJ and LC (Table 6A, west to east). In sharp contrast, both SJ and LC are massively allocated to the *opposite* MT from the same geographic provenance (Table 6B, both P-values < 0.0001). Similarly, AMOVAs show that a much larger proportion of genetic variation is explained by *Lake* (17.4%, P < 0.001) than by *MT* (i.e., SJ vs. LC; 0%, P = 0.846,Table 5D) (ΔAIC = 75.0). This result is observed regardless of the subset of lakes with SJ considered.

INLAND LAKES – SHORTJAW CISCO VS. SYMPATRIC MORPHOTYPES

White Partridge Lake

Evidence for a distinct form corresponding to SJ is equivocal for White Partridge Lake. Estimates from STRUCTURE yield K=2 clusters, but these have no biological meaning (not shown); FLOCK finds no evidence for more than one cluster (Table 3). However, when examining the clustering solution for K = 3 by STUCTURE (Figure 2H), it is apparent that one cluster comprises most large individuals with low gill raker number (LGRI in Table 1). The two other

MTs (small fish with high (LC) or low (LGRs) gill raker number) form the bulk of the second cluster. A third 'shadow cluster' comprises 4-5 individuals. Structure *q*-values are plotted against gill raker number and body size for all three MTs in Figures 4B and 4C. Very high *q*-values ($q > 0.8$) seem more frequent among putative SJ individuals (LGRI), but individuals span the full range of *q*-values. Similarly, individuals of the two other MTs tend to have lower but highly variable q-values. The high variability of *q*-values for each MT is also apparent in Figure 4C. Nonetheless, LGRI form a cohesive genetic group: the re-allocation matrix shows that 77% of the LGRIs are correctly re-allocated to this MT while the other two MTs are rarely re-allocated to LGRI (Table 4E). LGRs and LC are nearly equally re-allocated to one another, suggesting no clear distinction between them. Overall, it seems that the putative SJ (LGRI) may be genetically distinct but that its phenotypic distinctiveness is not associated with gill raker counts.

Trout Lake

In Trout Lake, there is clear evidence for two clusters, both with STRUCTURE and FLOCK (Table 3, Figure 2I). Clusters very closely correspond to MTs, with all but one SJ individual clearly belonging to the same cluster. A few LC also have strong ancestry in this cluster. Figures 4D and 4E show that fish with high *q*-values tend to have fewer gill rakers but can be of any size. Most correspond to fish labelled as SJ. In contrast, fish with low *q*-values were all small and tended to have more than 39 gill rakers. However, a few fish labelled as LC were genetically more similar to SJ. Finally, several small fish with mid- to high gill raker number had intermediate *q*-values (0.2-0.8). The re-allocation matrix is nearly perfect (Table 4F).

Brule Lake

In Brule Lake there is no evidence for two clusters with either clustering program (Table 3, Figure 2J). The re-allocation matrix suggests that SJ can be re-allocated to themselves (75%), but LC are also more frequently re-allocated to SJ (7/11) than to LC (4/11) (Table 4G). It is worth reporting that there was particularly little genetic polymorphism in both MTs from Brule Lake. The number of detected AFLP peaks (or loci) for each fish was typical of the overall dataset (mean of 70 in Brule vs. overall mean of 72 peaks), but the proportion of variable loci among fish was very low (10% vs. 35-40%).

Lake of the Woods

In Lake of the Woods, there is no evidence for two clusters with either method (Table 3, Figure 2K), possibly because there were only 8 SJ samples (with 30 LC samples). In the re-allocation matrix, all but one fish are re-allocated to the LC group (Table 4H).

Lake Athapapuskow

In Lake Athapapuskow, there is evidence supporting the genetic distinctiveness of SJ. STRUCTURE defines three clusters while FLOCK identifies only two genetic clusters (Table 3, Figure 2L). However, Fig. 2L shows that there are really only two clusters present, with a third 'shadow' cluster being formed by Structure at K=3. One cluster is mostly composed of SJ individuals, while the other is mostly composed of LC. In this lake, fish with 36 or less gill rakers were labelled as SJ (Murray 2006), the others were called LC. Most SJ had high *q*-values ($q > 0.7$), but six fish had values more typical of the other cluster ($q < 0.4$). Among LC, *q*-values were all in the 0–0.5 range (Figure 4F).The re-allocation matrix is highly significant, with 75% of LC and 88% of SJ correctly re-allocated (Table 4I). Although some samples are small, having both MTs available from different lake sectors allowed for a nested AMOVA (Table 5E). Genetic variation is better explained by *MT* (4.95%, but $P = 0.097$) than by *Sector* (0%, $P = 0.747$), even though the former model does not have better explanatory power ($\Delta AIC = 0.5$).

Great Slave Lake

In Great Slave Lake, it appears that the 'adfluvial' ciscoes are slightly differentiated, but SJ are not. Four MTs, and another species (Least Cisco, *C. sardinella*) were reported in this lake (Muir et al. 2013). Excluding *C. sardinella* (easily identifiable with AFLPs), there was no evidence for more than one cluster with either method (Table 3, Figure 2M). The re-allocation matrix to the four MTs, however, indicates that the river spawning cisco (RS, i.e., 'adfluvial' in Muir et al. 2013) may be genetically distinct. Indeed, this is the only MT for which a large proportion of specimens are correctly re-allocated (79%, Table 4J). In an AMOVA (Table 5F), a significant proportion of the genetic variation is attributed to *MT* (3.6%, P = 0.009), contrary to *Sector* (0%, P = 0.497), but both models did not differ in explanatory power (ΔAIC = 0.9).

Great Bear Lake

In Great Bear Lake, there is evidence for a SJ-like MT (with few gill rakers) inhabiting deeper waters (Howland et al., unpublished Res Doc). Three clusters are defined by STRUCTURE, but only two with FLOCK (Table 3, Figure 2N, Figure 3). With both methods, one cluster comprises most fish from the deep sectors in both Keith Arm and Dease Arm. All but a few fish from deep habitats had high *q*-values (Figure 4G). Fish from shallow sectors had lower but more variable *q*-values, especially in Dease Arm. In Dease Arm, fish from deep sets had lower gill raker number and most had high *q*-values (Figure 4H). A similar pattern is observed in Keith Arm although there is more overlap in gill raker number between fish with high vs. low *q*-values (Figure 4I). Samples tended to re-allocate to other samples from the same depth habitat when misclassified within sector (Table 4K). This trend is confirmed by nearly perfect re-allocation when only depth habitats are used to form reference groups (Table 4L). The AMOVA using *Habitat* (i.e., depth) vs. *Arm* (lake sector) as main factor explains more variation, but this proportion is not significant (Table 5G).

DISCUSSION

This study presents evidence for the genetic distinctiveness of a morphotype commonly recognized as the Shortjaw Cisco, and it supports the parallel and repeated origin of this morphotype in several locations throughout its Canadian range. The Plasticity Hypothesis is clearly dismissed. Within many lakes, fish that are morphologically recognized and designated by experts as Shortjaw Cisco are genetically distinct from other sympatric cisco morphotypes (Table 7). This is clearly the case in Lake Nipigon, Trout Lake, and Lake Athapapuskow. Evidence for SJ being distinct is also found for Great Bear Lake but the phenotypic attributes of SJ may be variable in the different arms of the lake. In White Partridge Lake, a distinct genetic group exists but its phenotypic attributes do not correspond to those which usually define SJ (number of gill rakers). In Lake Superior, there is weak evidence for SJ being distinct in only one sector of the lake. In Great Slave Lake, there was no evidence that SJ was genetically distinct. This result parallels the lack of phenotypic differentiation of this morphotype found when Muir et al. (2013) analysed cisco morphology without any *a priori* classification. In Lake of the Woods, the lack of evidence for SJ being distinct may be linked to low sample size, while the unusually low level of genetic diversity probably impedes the detection of genetic differentiation in Brule Lake. Overall, it appears that genetic markers will not always allow diagnosing individual fish as SJ in all lakes, and this may be impossible in some lakes. Nevertheless, the overall evidence is strongly in favor of recognizing fish with low GR (and possessing other attributes deemed typical of SJ on the basis of careful morphological investigations) as distinct units.

The Good Species Hypothesis is not supported for the specific case of the Shortjaw Cisco. Using the entire dataset, the highest level of structuring defines 'eastern' vs. 'western' clusters of lakes. Each cluster comprises SJ, LC, and other MTs. Thus, SJ (and other MTs) is more

closely related to other MTs from its regional clusters than to SJ from other regions. This pattern is incompatible with SJ forming a distinct lineage. Rather, it suggests that the genetic composition of a fish is first and foremost defined by where it is found rather than by its phenotypic attributes. This is further exemplified by the successful allocation of individuals across a taxonomic boundary vs. the impossibility of allocating fish across a geographic boundary. It is worth noting that this interpretation applies to both the Great Lakes (at least Superior and Nipigon) and inland Canadian lakes. Indeed, lakes Nipigon and Superior belong to different clusters (west and east, respectively) indicating that SJ (and likely the other MTs) does not form a distinct lineage in the Great Lakes.

Results are in line with the predictions derived from the Parallel Origin Hypothesis (Table 4). Evidence for a distinct SJ genetic entity is found only at small spatial scales, i.e., within (some) lakes (see above and Table 7). Hierarchical clustering (Figure 3), AMOVAs (Table 5D) and re-allocation (Table 6) concur in supporting that SJ are more closely related to sympatric forms than to other allopatric SJ. Again, the subdivision of lakes into eastern and western groups strongly indicates that there are at least two parallel origins of the SJ. Given that lakes Nipigon and Superior belong to different clusters, it strongly suggests that SJ was independently derived in each of these two Great Lakes. Unfortunately, data is too scarce in Lake Huron, with only two specimens available.

The degree of genetic distinctiveness between SJ and LC (and other MTs) was highly variable across lakes. The terminology of Hendry (2009) is also useful to describe how ciscoes have proceeded to different, but partially overlapping stages along the continuum of ecological speciation. For example, discontinuous phenotypic variation may be associated with weak or intractable genetic evidence for reproductive isolation in some lakes (e.g., Great Slave Lake, state 2) while differences parallel clear, but possibly reversible, reproductive isolation in others (i.e., Trout Lake, state 3 or 4 in Hendry 2009). These differences in the degree of genetic distinctiveness may depend on a variety of factors that are difficult to disentangle. If SJ indeed arose independently in many postglacial lakes, the age of diverging lineages should be roughly equivalent, so time since divergence is an unlikely explanation. The incomplete action of genetic drift in larger populations from larger lakes could be invoked for Lake Superior and Great Slave Lake. Yet again, drift has proceeded far enough in other large lakes to generate a clear signal of differentiation (e.g., Lake Nipigon, Great Bear Lake). The scope of ecological opportunities and the degree of differentiation of niches or habitat may also be at play, but this study is not informative on these matters. Finally, the lack of differentiation could also be ascribed to ecological perturbations favouring hybridization of young lineages. We had anticipated that this would diminish the distinctiveness of SJ in lakes with a history of ecological perturbations (i.e., the Great Lakes proper) vs. pristine inland lakes such as Great Slave Lake or Great Bear Lake. Unfortunately, our results do not allow a conclusion to be reached on this possibility. First, the level of SJ differentiation contrasts between the two relatively pristine lakes, being stronger in Great Bear Lake and undetectable in Great Slave Lake. Second, having results from a single 'true' Great Lake (Lake Superior) in the eastern cluster provides a poor comparison. Nevertheless, the presence of two clusters loosely associated with MTs within Lake Superior (Figure 2D,G), coupled with the weak evidence for the genetic distinctiveness of SJ (and Kiyi) in one sector of the lake could be the result of recent hybridization. This speculative interpretation clearly needs further investigation.

To conclude, it is worth noting that the results of this study are in complete agreement with previous work. Turgeon et al. (1999) also found evidence for SJ in Lake Nipigon. Turgeon and Bernatchez (2001a,b) found evidence for the clinal mixing of two genetic groups labelled as glacial races. As shown in the current study, Lake Nipigon and lakes located west of it formed one group of lakes that were characterized by the higher representation of one mtDNA glacial

race. Analyses of sympatric pairs also favoured the parallel, likely sympatric origin of ciscoe morphotypes (Turgeon and Bernatchez 2003). It is worth noting that these highly similar conclusions, at various spatial scales, are reached with two independent datasets using different genetic markers (7 microsatellites vs. 290 AFLP loci) on totally different samples from an overlapping but different set of morphotypes and lakes.

ACKNOWLEDGEMENTS

This research was funded by a Great Lake Fishery Commission grant (Fishery Research Program) entitled 'Sympatric speciation and patterns of morphological and genetic diversity among ciscoes in deepwater lakes" held by Scott Reid, Julie Turgeon, and Tom Pratt.

REFERENCES CITED

Bernatchez, L. 2004. Ecological theory of adaptive radiation: an empirical assessment from Corigonine fishes (Salmoniformes). *In* Evolution Illuminated: Salmon and Their Relatives. Edited by A.P.H.a.S. Stearns. Oxford University Press, Oxford, U.K. p. 176-207.

Bonin, A., Bellemain, E., Eidesen, P.B., Pompanon, F., Brochmann, C., and Taberlet, P. 2004. How to track and assess genotyping errors in population genetics studies. Mol. Ecol. 13: 3261-3273.

Clarke, R.M. 1973. The systematics of ciscoes (Coregonidae) in central Canada. Thesis (Ph.D.), Department of Zoology, University of Manitoba, Winnipeg, MB.

Duchesne, P., and Bernatchez, L. 2002. AFLPOP: a computer program for simulated and real population allocation, based on AFLP data. Mol. Ecol. Notes 2: 380-383.

Duchesne, P., and Turgeon, J. 2009. FLOCK: a method for quick mapping of admixture without source samples. Mol. Ecol. Resour. 9: 1333-1344.

Duchesne, P., and Turgeon, J. 2012. FLOCK provides reliable solutions to the "number of populations" problem. J. Hered. 103: 734-743.

Evanno, G., Regnaut, S., and Goudet, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol. Ecol. 14: 2611-2620.

Excoffier, L., and Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10: 564-567.

Falush, D., Stephens, M., and Pritchard, J.K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. Mol. Ecol. Notes 7: 574-578.

Halverson, K., Heard, S.B., Nason, J.D., and Stireman, J.O. 2008. Origins, distribution, and local co-occurrence of polyploid cytotypes in Solidago altissima (Asteraceae). Am. J. Bot. 95: 50-58.

Hendry, A.P. 2009. Ecological speciation! Or the lack thereof? Can. J. Fish. Aquat. Sci. 66: 1383-1398.

Howland, K., Gallagher, C., Boguski, D., Reist, J., Chavarie, L., and Wiley, S. 2013. Variation in morphology, life history and ecology of cisco in Great Bear Lake, Northwest Territories, Canada. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/nnn. vi + xx p.

Hubisz, M.J., Falush, D., Stephens, M., and Pritchard, J.K. 2009. Inferring weak population structure with the assistance of sample group information. Mol. Ecol. Resour. 9: 1322-1332.

Hudson, A.G., Vonlanthen, P., Müller, R., and Seehausen, O. 2007. The geography of speciation and adaptive radiation in coregonines. Adv. Limnol. 60: 111-146.

Koelz, W. 1929. Coregonid fishes of the Great Lakes. Bull. U.S. Bur. Fish. 43: 297-643.

Muir, A.M., Vecsei, P., Power, M., Krueger, C.C., and Reist, J.D. 2013. Morphology and life history of the Great Slave Lake ciscoes (Salmoniformes: Coregonidae). DFO Can. Sci. Advis. Sec. Res. Doc. 2013/nnn. vi + 30 p.

Murray, L. 2006. A morphological examination of sympatric cisco forms in four lakes with specific reference to the occurrence of shortjaw cisco (*Coregonus zenithicus*) in Manitoba. Thesis (M.Sc.), Department of Zoology, University of Manitoba, Winnipeg, MB.

Pratt, T.C. 2012. The distribution and abundance of deepwater ciscoes in Canadian waters of Lake Superior. *In* Biology and Management of Coregonid Fishes. Edited by R.F. Tallman, K.L. Howland, M.D. Rennie and K.H. Mills. p. 25-41.

Pratt, T.C. 2013. The biology and relative abundance of Shortjaw Cisco (*Coregonus zenithicus*) in Lake Nipigon. DFO Can. Sci. Advis. Sec. Res. Doc. 2013/nnn. iv + 13 p.

Pratt, T.C., and Chong, S.C. 2012. Contemporary life history characteristics of Lake Superior deepwater ciscoes. Aquat. Ecosyst. Health Manag. 15: 322-332.

Pritchard, J.K., Stephens, M., and Donnelly, P. 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Rosenberg, N.A. 2004. DISTRUCT: a program for the graphical display of population structure. Mol. Ecol. Notes 4: 137-138

Rundle, H.D., Nagel, L., Wenrick Boughman, J., and Schluter, D. 2000. Natural selection and parallel speciation in sympatric sticklebacks. Science 287: 306-308.

Schluter, D. 1996. Ecological speciation in postglacial fishes. Phil. Trans. R. Soc. B 351: 807-814.

Scott, W.B., and Crossman, E.J. 1973. Freshwater Fishes of Canada. Bull. Fish. Res. Bd. Can. 184: 1092 p.

Smith, G.R., and Todd, T.N. 1984. Evolution of species flocks of fishes in north temperate lakes. *In* Evolution of Fish Species Flocks. Edited by A.A. Echelle and I. Kornfield. University of Maine at Orono Press, Orone. p. 45-68.

Taylor, E.B., and McPhail, J.D. 2000. Historical contingency and ecological determinism interact to prime speciation in sticklebacks, *Gasterosteus*. Proc. R. Soc. B 267: 2375-2384.

Todd, T.N., and Stedman, R.M. 1989. Hybridization of ciscoes (*Coregonus* spp.) in Lake Huron. Can. J. Zool. 67: 1679-1685.

Turgeon, J., and Bernatchez, L. 2001a. Clinal variation at microsatellite markers reveals extensive historical secondary contacts between refugial races of *Coregonus artedi* (Teleostei: Coregoninae). Evolution 55: 2274-2286.

Turgeon, J., and Bernatchez, L. 2001b. Mitochondrial DNA phylogeography of lake cisco (*Coregonus artedi*): evidence supporting extensive secondary contacts between two glacial races. Mol. Ecol. 10: 987-1001.

Turgeon, J., and Bernatchez, L. 2003. Reticulate evolution and phenotypic diversity in North American ciscoes, *Coregonus* spp. (Teleostei: Salmonidae): implications for the conservation of an evolutionary legacy. Conserv. Genet. 4: 67-81.

Turgeon, J., Estoup, A., and Bernatchez, L. 1999. Species flocks in the North American Great Lakes: molecular ecology of Lake Nipigon ciscoes (Teleostei: Coregonidae: Coregonus). Evolution 53: 1857-1871.

Table 1. Sample sizes for each lake (and sector within lakes where applicable). Ciscoes are classified using common names used by fish biologists providing samples (LC: Lake Cisco, SJ: Shortjaw Cisco, BF: Blackfin Cisco, KI: Kiyi, BL: Bloater, BE: "Big eye", RS: "River Spawning"; LGRl/s: "Low Gill Raker – large/small body size). In Great Bear Lake, fish from the deep sections of Keith Arm and Dease Arm are labelled as SJ-like morph (Howland et al. unpublished data). Samples included in the 'Shortjaw dataset' are underlined.

CODE - Lake/Sector	Common Name/Morphotype									Total
	LC	SJ	LGRl	LGRs	BF	KI	BL	BE	RS	
HUR - Lake Huron[5]	61	2					37			100
Main Basin	*30*						*16*			*46*
North Channel	*31*	*2*					*21*			*54*
MCH - Lake Michigan[5]							27			27
NIP - Lake Nipigon[5]	61	60			63		62			246
Southeast	*31*	*30*			*32*		*30*			*123*
Southwest	*30*	*30*			*31*		*32*			*123*
ONT - Lake Ontario[4]	31									31
SUP - Lake Superior[5]	57	64				60	62			243
East	*31*	*31*				*31*	*31*			*124*
Northwest	*26*	*33*				*29*	*31*			*119*
Great Lakes	**210**	**126**			**63**	**60**	**188**			**647**
BEA - Great Bear Lake[1]	76	49								125
Dease Arm	*30*	33								*63*
Keith Arm	*46*	*16*								*62*
SLV - Great Slave Lake[2]	11	16						16	28	71
Beaulieu									*13*	*13*
Christie Bay	*7*	*14*						*7*		*28*
Red Cliff Bluffs	*4*	*2*						*9*		*15*
Tartan Rapids									*15*	*15*
ATA - Lake Athapapuskow[3]	24	33								57
Net 2	*19*	*3*								*22*
Net 3	*5*	*26*								*31*
Net 4		*6*								*6*
CHU - Churchill River Estuary[1]	20									20
LOW - Lake of the Woods[4]	31	8								39
BIG - Biggar Lake[4]	27									27
BRU - Brule Lake[4]	11	32								43
GRA - Grand Lake[4]	32									32
KEN - Kennebec Lake[4]					32					32
MAN - Lake Manitou[4]	31									31
SCO - Scorch Lake[4]					28					28
SHA - Sharbot Lake[4]	29									29
SIM - Lake Simcoe[4]	32									32
TRO - Trout Lake[4]	36	31								67
WPT - White Partridge Lake[4]	31		30	30						91
Inland Lakes	**424**	**136**	**30**	**30**	**60**			**16**	**28**	**724**
Shortjaw Dataset	**338**	**293**	**30**	**30**						**691**
Grand Total	**601**	**295**	**30**	**30**	**123**	**60**	**188**	**16**	**28**	**1371**

Fish samples provided by 1: Kim Howland (DFO), 2: A. Muir (GLFC), 3: J.D Reist (DFO), 4: Scott Reid (OMNR), 5: Tom Pratt (DFO).

Table 2. AFLP analyses: Primer pairs with selective nucleotides, level of polymorphism, and co-migration strategy (Dye, Co-migration).

Primer pair code	Primer EcoRI +3	Primer MseI +4	Dye	N loci	N polymorphic loci (5% criterion)	Size Range (bp)	Co-migration
G5	AAC	CACT	NED	31	25	50 - 292	1
D8	ACC	CACA	6-FAM	71	54	59 - 370	1
F4	ATG	CATA	VIC	87	72	52 - 470	1
G2	AAC	CATC	NED	39	36	56 - 407	2
D3	ACC	CATG	6-FAM	71	46	53 - 384	2
F14	ATG	CAGC	VIC	73	57	52 - 470	2

Table 3. Number of genetic clusters (K) identified with structure and flock for all lakes, the Great Lakes, and single lakes where the Shortjaw Cisco was reported. For flock, only the top-level clustering is reported (See Figure 3 for lower levels).

			STRUCTURE		FLOCK
Dataset (Lake)	N ind.	K tested	Ln P(X|K)	Evanno*	K
Total	1371	1-20	2	15	2
Great Lakes	647	1-10	8	2	2
Nipigon	246	1-8	2	2,6	2
Superior	243	1-8	1	2	-
Huron	100	1-6	1	2	-
White Partridge	91	1-4	2	2	-
Trout	67	1-4	2	2	2
Brule	43	1-4	2	2	-
Lake of the Woods	39	1-4	1	2	-
Athapapuskow	59	1-6	3	3	2
Great Slave	71	1-8	2	2	-
Great Bear	125	1-4	3	3	2

-: There is no evidence that K > 1

* Evanno criterion leads to estimates of $K \geq 2$ (K = 1 is not possible)

Table 4. Re-allocation matrix, based on individual AFLP genotypes, of fish identified as belonging to a given morphotype (or sector, or habitat) to reference groups defined by the same criterion. Bold numbers on the diagonal indicate correctly allocated fish. See Table 1 for codes of morphotypes.

A) Great Lakes – Re-allocation to Lakes

	Number of Specimens Among				
Allocated to	Huron	Ontario	Michigan	Nipigon	Superior
Huron	**57**	1	12	0	47
Ontario	0	**30**	0	0	0
Michigan	9	0	**11**	0	9
Nipigon	1	0	0	**244**	5
Superior	33	0	4	2	**182**
% Correct	57%	97%	41%	99%	75%

B) Great Lakes – Re-allocation to Morphotypes

	Number of Specimens Among				
Allocated to	Bloater	Lake Cisco	Shortjaw	Blackfin	Kiyi
Bloater	**67**	45	13	6	4
Lake Cisco	44	**94**	13	1	8
Shortjaw	16	16	**55**	7	1
Blackfin	25	34	14	**49**	0
Kiyi	36	21	29	0	**47**
% Correct	36%	45%	44%	78%	78%

Table 4. continued.

C) Lake Nipigon – Re-allocation to Morphotypes

	Number of Specimens Among			
Allocated to	Blackfin	Bloater	Lake Cisco	Shortjaw
Blackfin	**38**	7	11	4
Bloater	9	**42**	15	1
Lake Cisco	12	10	**31**	4
Shortjaw	4	3	4	**51**
% Correct	60%	68%	51%	85%

D) Lake Superior – Re-allocation to Morphotype by Sector

Sector 1 - East				
	Number of Specimens Among			
Allocated to	Bloater	Kiyi	Lake Cisco	Shortjaw
Bloater	**11**	7	10	5
Kiyi	6	**9**	3	8
Lake Cisco	8	6	**12**	3
Shortjaw	6	9	6	**15**
% Correct	35%	29%	39%	48%

Sector 2 - Northwest				
	Number of Specimens Among			
Allocated to	Bloater	Kiyi	Lake Cisco	Shortjaw
Bloater	**12**	6	9	2
Kiyi	7	**18**	2	2
Lake Cisco	10	1	**11**	6
Shortjaw	2	4	4	**23**
% Correct	39%	62%	42%	70%

Table 4. continued.

E) White Partridge Lake

	Number of Specimens Among		
Allocated to	HGRs	LGRs	LGRI
HGRs	**13**	17	4
LGRs	17	**12**	3
LGRI	1	1	**23**
% Correct	42%	40%	77%

F) Lake Trout

	Number of Specimens Among	
Allocated to	Lake Cisco	Shortjaw
Lake Cisco	**33**	1
Shortjaw	3	**30**
% Correct	92%	97%

G) Brule Lake

	Number of Specimens Among	
Allocated to	Lake Cisco	Shortjaw
Lake Cisco	**4**	8
Shortjaw	7	**24**
% Correct	36%	75%

Table 4. continued.

H) Lake of the Woods

	Number of Specimens Among	
Allocated to	Lake Cisco	Shortjaw
Lake Cisco	**31**	7
Shortjaw	0	**1**
% Correct	100%	13%

I) Lake Athapapuskow

	Number of Specimens Among	
Allocated to	Lake Cisco	Shortjaw
Lake Cisco	**18**	**4**
Shortjaw	6	**29**
% Correct	75%	88%

J) Great Slave Lake

	Number of Specimens Among			
Allocated to	'Big Eye'	Lake Cisco	Riverine	Shortjaw
'Big Eye'	**7**	2	5	3
Lake Cisco	0	**2**	0	1
Riverine	7	5	**22**	3
Shortjaw	2	2	1	**9**
% Correct	44%	18%	79%	56%

Table 4. continued.

K) Great Bear Lake – By Sector and habitat – (Ciscoes from the deep and shallow areas are classified as Shortjaw and Lake cisco by Howland et al. 2013)

allocated to		number of specimens among:			
		Dease Arm		Keith Arm	
		Shallow	Deep	Shallow	Deep
Dease Arm	Shallow	**20**	3	0	1
	Deep	2	**24**	0	8
Keith Arm	Shallow	8	2	**46**	0
	Deep	0	4	0	**7**
% correct		67%	73%	100%	44%

L) Great Bear Lake – By Habitat (Ciscoes from the deep and shallow areas are classified as Shortjaw and Lake cisco by Howland et al. 2013).

	Number of Specimens Among	
Allocated to	Shallow	Deep
Shallow	**74**	6
Deep	2	**43**
% Correct	97%	88%

Table 5. Analyses of Molecular Variation (AMOVA) testing contrasting hierarchical models with Lake (or lake Sector) vs. Morphotype (MT) as the main factor.

A) Great Lakes

Lake nested within MT	df	SS	% Var.	Fixation index	P
Among MT	4	423	1.55	0.016	0.107
Among lakes within MT	18	1165	8.82	0.090	< 0.001
Within lakes	624	10875	89.63	0.104	< 0.001

MT nested within Lake	df	SS	% Var.	Fixation index	P
Among lakes	4	961	9.27	0.093	< 0.001
Among MT within lake	18	628	3.14	0.035	< 0.001
Within MT	624	10875	87.59	0.124	< 0.001

B) Lake Nipigon

Sector nested within MT	df	SS	% Var.	Fixation index	P
Between MT	3	208	3.43	0.034	0.006
Among sectors within MT	4	103	0.96	0.010	< 0.001
Within sectors	238	4687	95.61	0.044	< 0.001

MT nested within Sector	df	SS	% Var.	Fixation index	P
Between sectors	1	48	0.17	0.002	0.416
Among MT within sector	6	262	3.81	0.038	< 0.001
Within MT	238	4687	96.01	0.040	< 0.001

Table 5. continued.

C) Lake Superior

Sector nested within MT	df	SS	% Var.	Fixation index	P
Among MT	3	120	1.3	0.013	0.052
Between sectors within MT	4	107	2.08	0.021	< 0.001
Within sectors	235	3807	96.62	0.034	< 0.001

MT nested within Sector	df	SS	% Var.	Fixation index	P
Between sectors	1	50	1.02	0.010	0.080
Among MT within sectors	6	177	2.6	0.026	< 0.001
Within MT	235	3807	96.37	0.036	< 0.001

D) All Lakes with Shortjaw Cisco and Lake Cisco

Lake nested within MT	df	SS	% Var.	Fixation index	P
Between MT	1	118	-3.02	-0.030	0.847
Among lakes within MT	8	1952	22.45	0.218	< 0.001
Within lakes	453	8160	80.58	0.194	< 0.001

MT nested within Lake	df	SS	% Var.	Fixation index	P
Among lakes	4	1735	17.36	0.174	< 0.001
Between MT within lake	5	334	4.75	0.057	< 0.001
Within MT	453	8160	77.89	0.221	< 0.001

Table 5. continued.

E) Lake Athapapuskow

Sectors nested within MT	df	SS	% Var.	Fixation index	P
Between MT	1	52	4.95	0.050	0.097
Among sectors within MT	3	63	0.33	0.004	0.340
Within sectors	52	1066	94.72	0.053	< 0.001

MT nested within sectors	df	SS	% Var.	Fixation index	P
Among sectors	2	61	-0.8	-0.008	0.747
Between MT within sectors	2	54	4.57	0.045	0.003
Within MT	52	1066	96.23	0.038	< 0.001

F) Great Slave Lake

Sectors nested within MT	df	SS	% Var.	Fixation index	P
Among MT	3	99	3.61	0.036	0.009
Among sectors within MT	4	80	-0.92	-0.010	0.784
Within sectors	63	1359	97.31	0.027	0.001

MT nested within sectors	df	SS	% Var.	Fixation index	P
Among sectors	3	80	-0.12	-0.001	0.497
Between MT within sectors	4	99	2.27	0.023	0.019
Within MT	63	1359	97.85	0.021	< 0.001

Table 5. continued.

G) Great Bear Lake (sectors are Dease Arm and Keith Arm; fish captured in deep waters are considered as SJ, otherwise as LC, see Howland et al. 2013)

Sector nested within Depth	d.f.	SS	% Var.	Fixation index	P
Between MT	1	94	4.21	0.042	0.338
Between sectors within MT	2	77	3.06	0.032	< 0.001
Within Sectors	121	2389	92.73	0.073	< 0.001

Depth nested within Sector	d.f.	SS	% Var.	Fixation index	P
Habitat nested within Arm					
Between Sectors	1	52	-1.4	-0.014	0.660
Between MT within Sector	2	119	6.93	0.068	< 0.001
Within MT	121	2389	94.48	0.055	< 0.001

Table 6. Allocation matrices of SJ and LC sampled in lakes from western and eastern locations (see Figure 2) across A) geographic (P = 0.736) and B) taxonomic boundaries (P <0001). The allocation is based on AFLP genotypes at 290 loci.

A) Allocation of MTs Across Geographic Boundary

East to West		
	Number of Specimens Among	
Allocated to	EAST_LC	EAST_SJ
WEST_LC	122	140
WEST_SJ	13	17

West to East		
	Number of Specimens Among	
Allocated to	WEST_LC	WEST_SJ
EAST_LC	77	79
EAST_SJ	50	38

Table 6. continued.

B) Allocation of location across taxonomic boundary

Lake Cisco to Shortjaw Cisco		
	Number of Specimens Among	
Allocated to	WEST_LC	EAST_LC
WEST_SJ	127	3
EAST_SJ	0	132

Shortjaw Cisco to Lake Cisco		
	Number of Specimens Among	
Allocated to	WEST_SJ	EAST_SJ
WEST_LC	117	2
EAST_LC	0	155

Table 7. Overall assessment of the evidence in favor of Shortjaw Cisco being genetically distinct from Lake Cisco (and other morphotypes). This includes the evidence for the most likely number of genetic clusters (K clusters, see Figure 2), the degree of correspondence with phenotypic attributes (q-phenotype, see Figure 4), the results of individual re-allocation to morphotypes (see Table 4), and evidence that phenotypes explain genetic variation better than lake sector (see Table 5). '-' indicates that the analyses were not performed.

Lake with Shortjaw	Analyses Performed				Evidence for Genetically Distinct Shortjaw
	K clusters	*q*-phenotype	Re-allocation	AMOVA	
NIP - Lake Nipigon	2	**Strong (DF1)**	**strong**	**strong**	**strong**
SUP - Lake Superior	1	-	weak	weak	weak
HUR - Lake Huron	1	-	-	-	-
WPT - White Partridge Lake	2	Weak (GR, size)	**strong**	-	some
TRO - Trout Lake	2	**Strong (GR, size)**	**strong**	-	**strong**
BRU - Brule Lake	1	-	none	-	none
LOW - Lake of the Woods	1	-	none	-	none
ATA - Lake Athapapuskow	2	**Strong (GR)**	**strong**	**strong**	**strong**
SLV - Great Slave Lake	1	-	none	weak*	none
BEA - Great Bear Lake	2 - 3	**Strong (depth, GR)**	**strong**	none	**Strong**

*: See text: significant component of genetic variation likely due to the differentiation of the river-spawning morph (not the SJ).

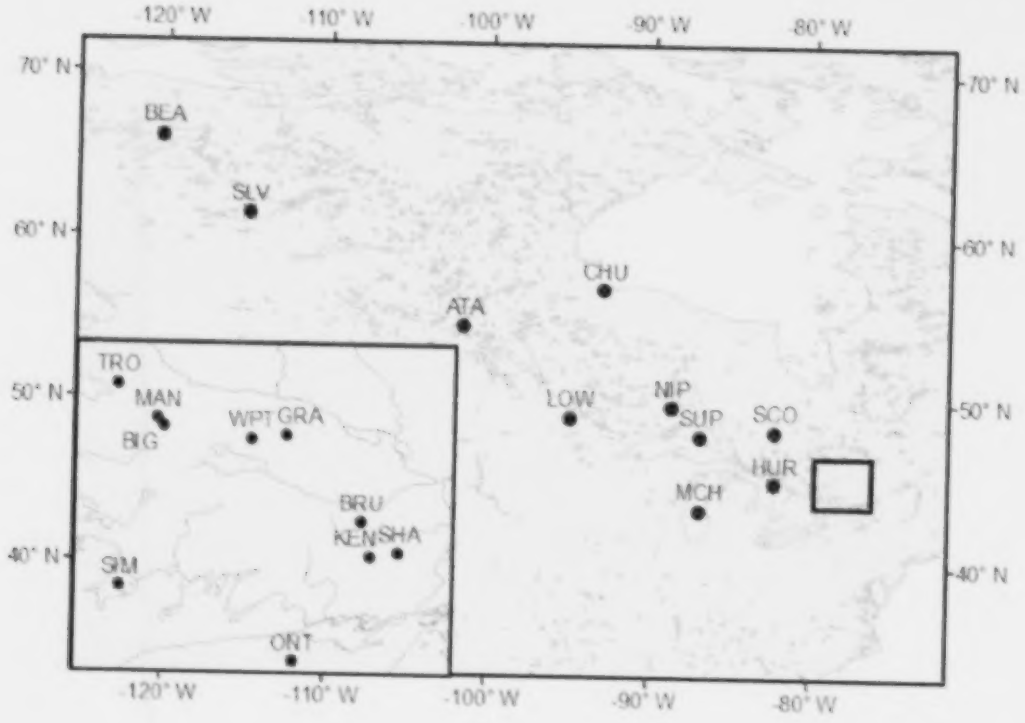
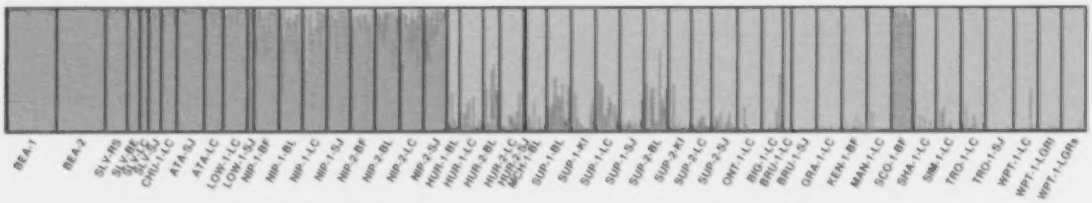
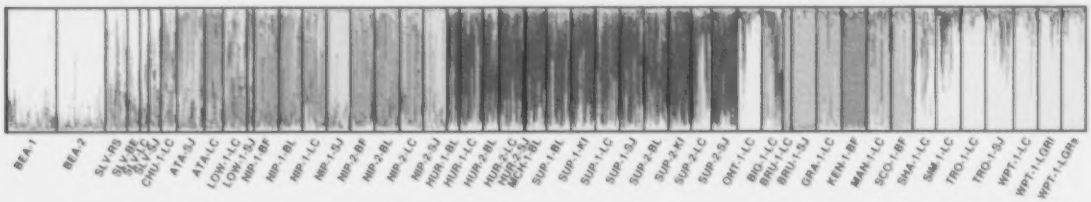


Figure 1. Map showing sampling location for ciscoes genetically characterized with AFLP markers. Codes are as per Table 1.

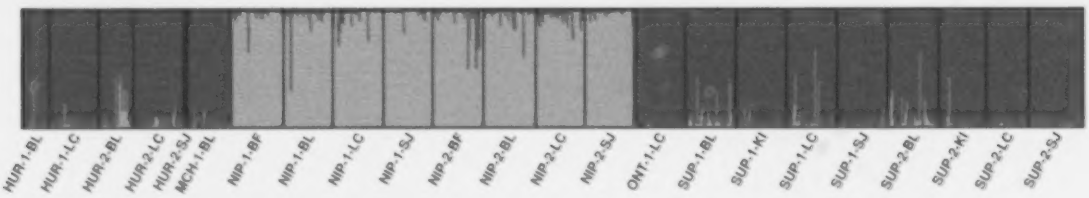
A) Entire Dataset, K=2 (Pritchard's criteria)



B) Entire Dataset, K=15 (Evanno's criteria)



C) Great Lakes Dataset, K=2 (Evanno's criteria)



D) Great Lakes samples, K=8 (Pritchard's criteria)

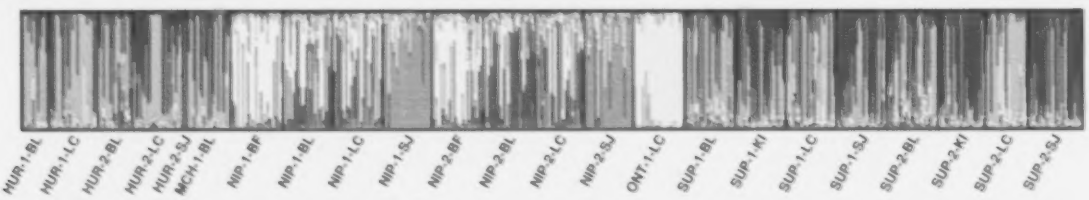
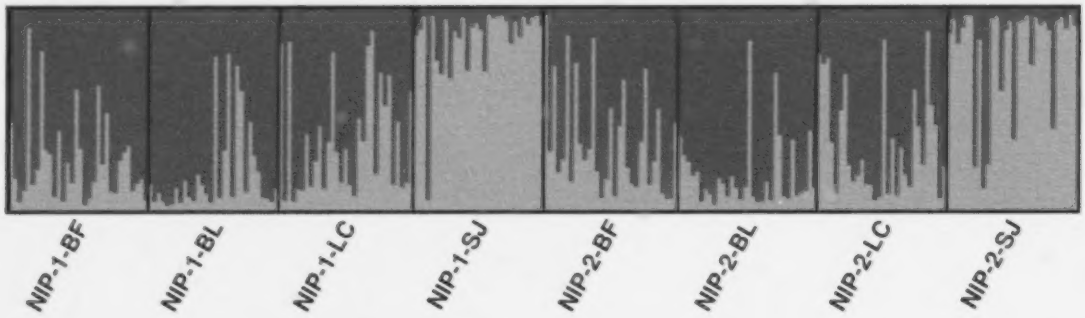
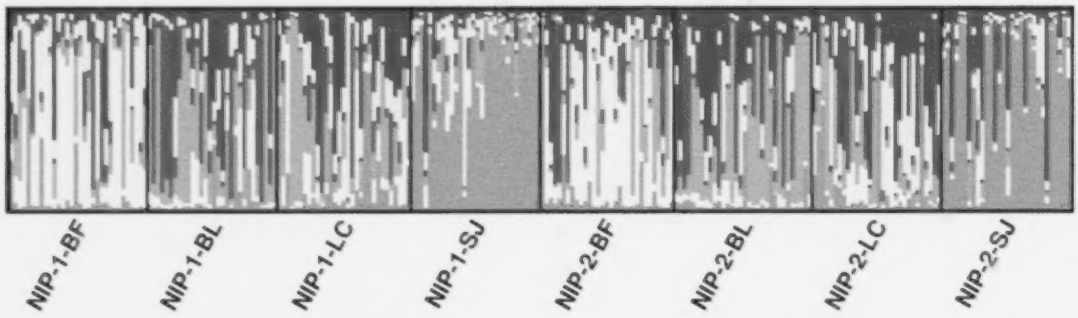


Figure 2(A-D). Ancestry of individual fish, grouped by samples (as per Table 1), in the most (or second most) likely number of K genetic clusters formed by STRUCTURE. In each panel, the vertical axis represents the coefficient of ancestry in each cluster (q-value), and each cluster is represented by a different color. Each vertical bar represents an individual fish.

E) Lake Nipigon, all samples, K=2 (preferred)



F) Lake Nipigon, K = 6 (second most likely solution)



G) Lake Superior, K=2 (K=1 is preferred)

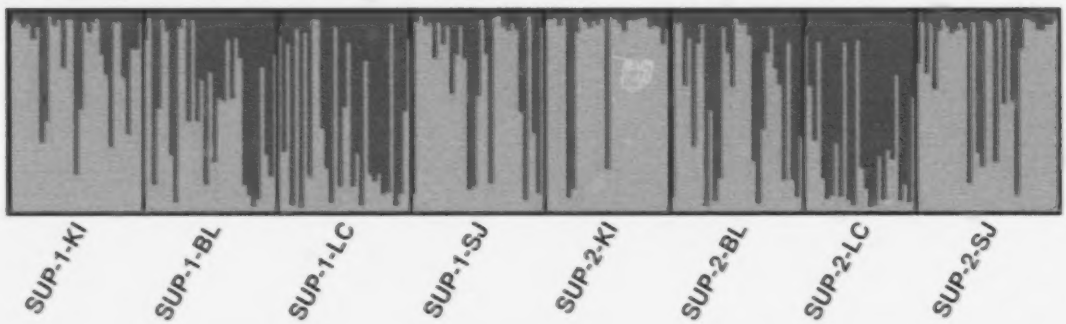
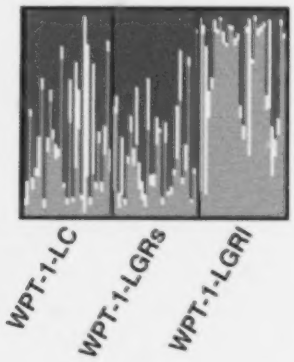
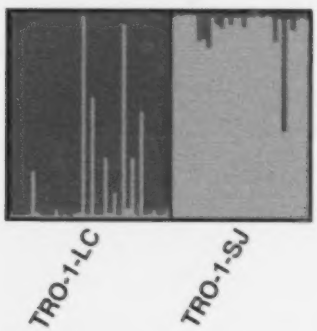


Figure 2(E-G). Ancestry of individual fish, grouped by samples (as per Table 1), in the most (or second most) likely number of K genetic clusters formed by STRUCTURE. In each panel, the vertical axis represents the coefficient of ancestry in each cluster (q-value), and each cluster is represented by a different color. Each vertical bar represents an individual fish.

H) White Partridge Lake, K= 3 (K=2 is preferred)



I) Trout Lake, K= 2



J) Brule Lake, K =2

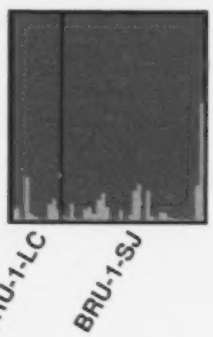
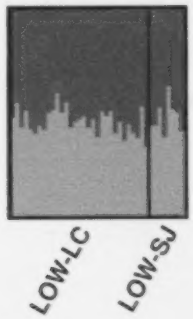
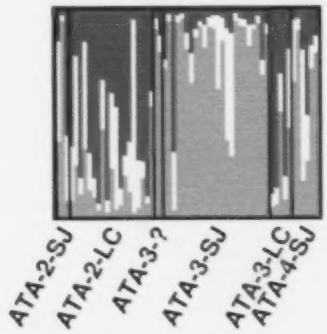


Figure 2(H-J). Ancestry of individual fish, grouped by samples (as per Table 1), in the most (or second most) likely number of K genetic clusters formed by STRUCTURE. In each panel, the vertical axis represents the coefficient of ancestry in each cluster (q-value), and each cluster is represented by a different color. Each vertical bar represents an individual fish.

K) Lake of the Woods, K= 2 (K=1 is preferred)



L) Lake Athapapuskow, K = 3



M) Great Slave Lake, K=2

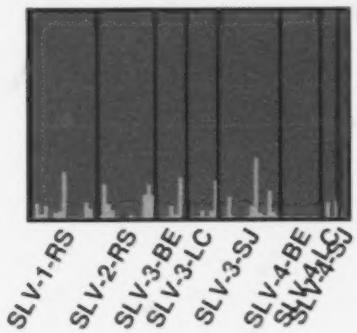


Figure 2(K-M). Ancestry of individual fish, grouped by samples (as per Table 1), in the most (or second most) likely number of K genetic clusters formed by STRUCTURE. In each panel, the vertical axis represents the coefficient of ancestry in each cluster (q-value), and each cluster is represented by a different color. Each vertical bar represents an individual fish.

N) Great Bear Lake, K=3. Samples are grouped by lake sector (Dease (D) and Keith (K) arms) and depth of sampling within each sector. Ciscoes from the deep and shallow areas are classified as Shortjaw and Lake cisco by Howland et al. 2013.

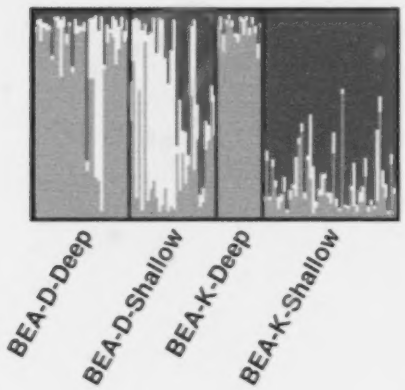


Figure 2(N). Ancestry of individual fish, grouped by samples (as per Table 1), in the most (or second most) likely number of K genetic clusters formed by STRUCTURE. The vertical axis represents the coefficient of ancestry in each cluster (q-value), and each cluster is represented by a different color. Each vertical bar represents an individual fish.

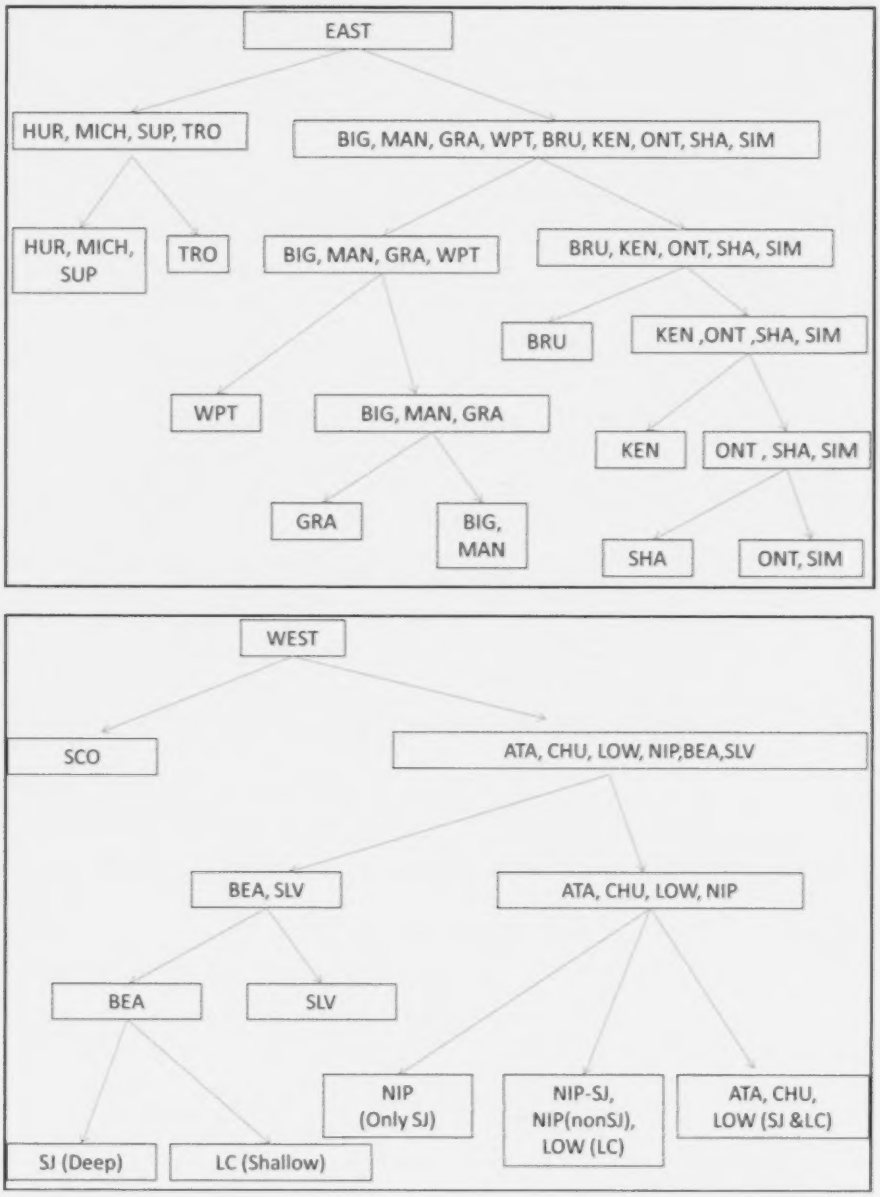
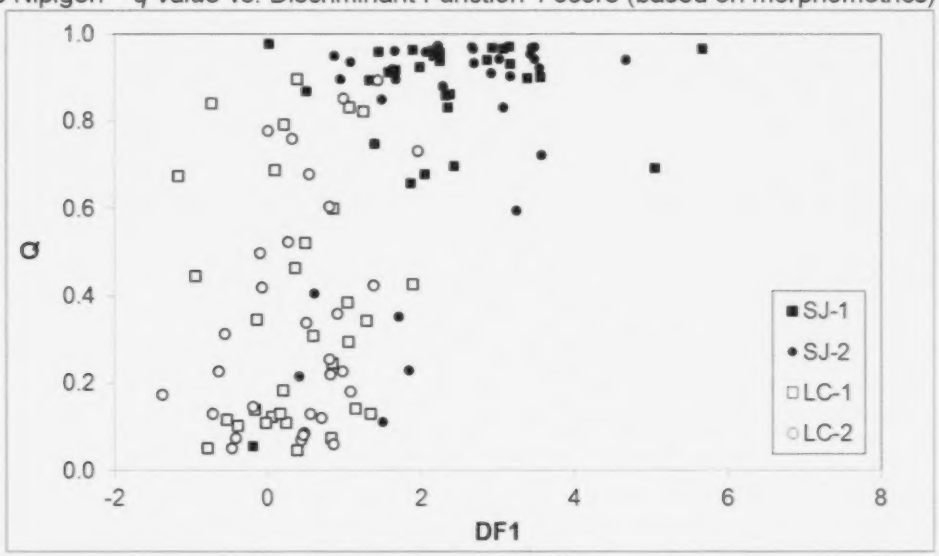


Figure 3. Hierarchical clusters formed by FLOCK using all samples. The first division groups lakes east of Nipigon ('East') vs. Nipigon and all lakes west of it ('West'). The sole exception is Scorch Lake (see text). Morphotypes in each lake are pooled when more than 90% of the fish from a lake belong to the same cluster. Codes are as per Table 1.

A) Lake Nipigon – *q-value* vs. Discriminant Function 1 score (based on morphometrics)



B) White Partridge Lake – *q-value* vs. Total Gill Raker number

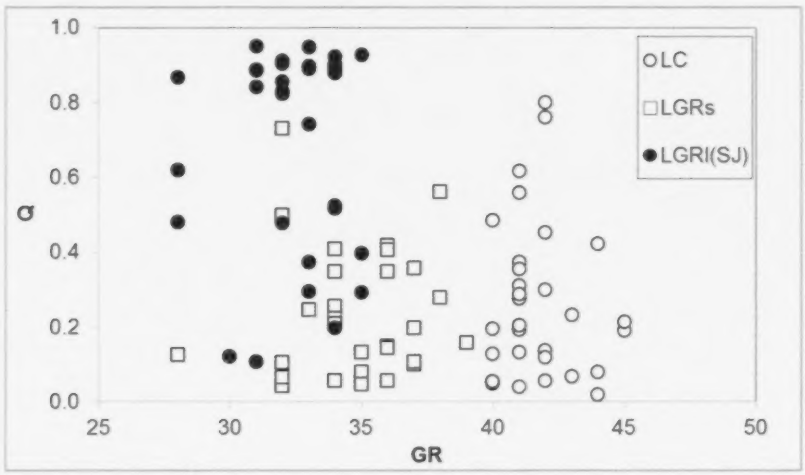
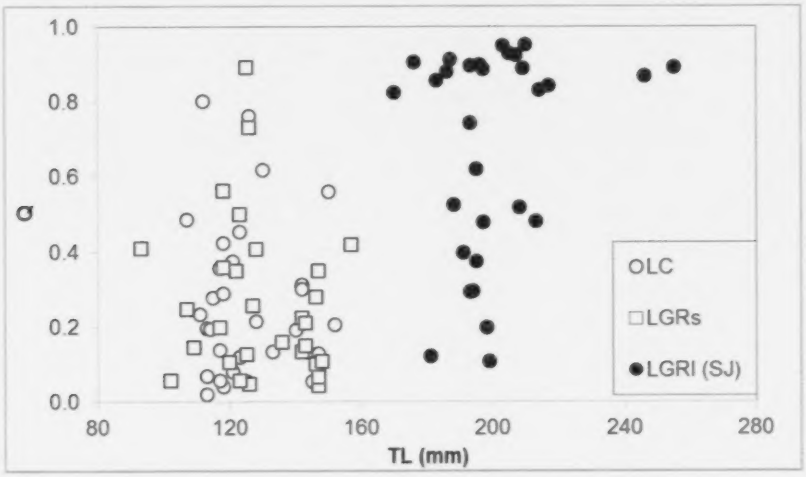


Figure 4(A-B). Association between individual genetic characteristics (q-value determined by Structure, expressed as the coefficient of ancestry of a genotype in the cluster best associated with SJ) and phenotypic attributes (DF1: Discriminant Function 1 score, GR: Total number of gill rakers, TL: Total body length in mm). Phenotypic attributes were from Murray (2006) or provided by researchers (S. Reid, OMNR; T. Pratt, DFO; K. Howland, DFO).

C) White Partridge Lake – *q-value* vs. Total Length (mm)



D) Trout Lake – *q-value* vs. Total Gill Raker number

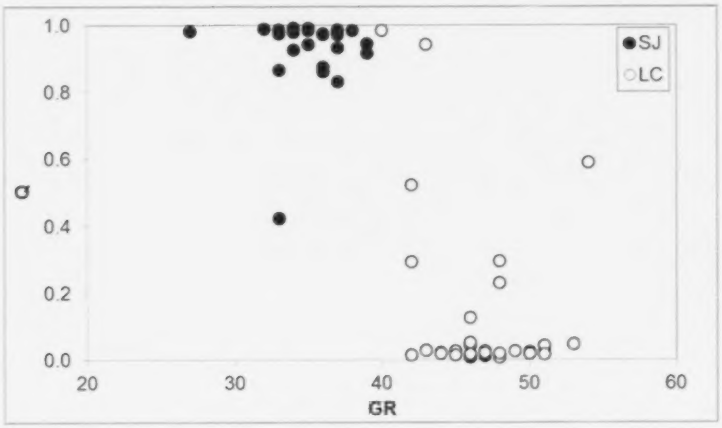
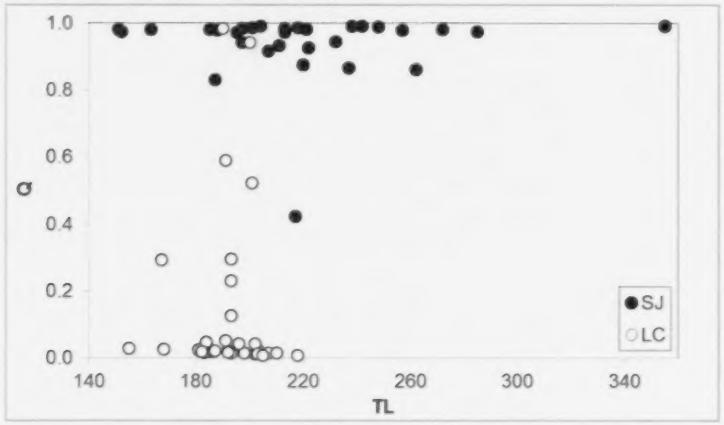


Figure 4(C-D). Association between individual genetic characteristics (q-value determined by Structure, expressed as the coefficient of ancestry of a genotype in the cluster best associated with SJ) and phenotypic attributes (DF1: Discriminant Function 1 score, GR: Total number of gill rakers, TL: Total body length in mm). Phenotypic attributes were from Murray (2006) or provided by researchers (S. Reid, OMNR; T. Pratt, DFO; K. Howland, DFO).

E) Trout Lake – *q-value* vs. Total length (mm)



F) Lake Athapapuskow – *q-value* vs. Total Gill Raker number

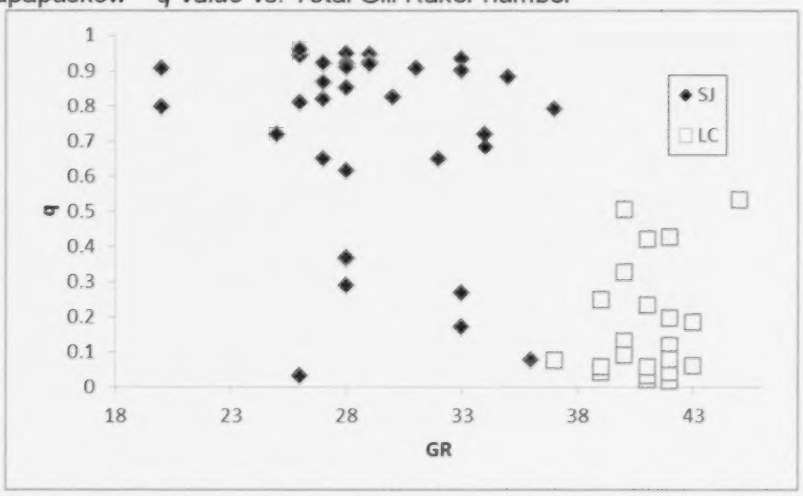
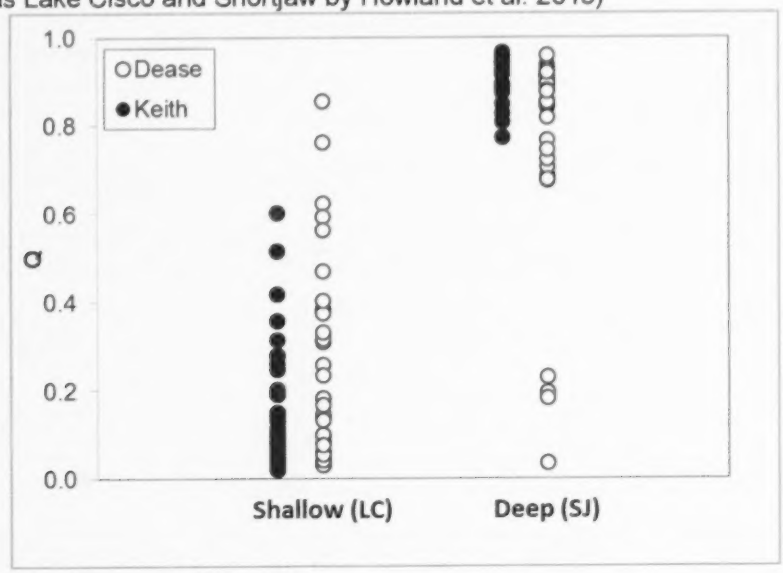


Figure 4(E-F). Association between individual genetic characteristics (q-value determined by Structure, expressed as the coefficient of ancestry of a genotype in the cluster best associated with SJ) and phenotypic attributes (DF1: Discriminant Function 1 score, GR: Total number of gill rakers, TL: Total body length in mm). Phenotypic attributes were from Murray (2006) or provided by researchers (S. Reid, OMNR; T. Pratt, DFO; K. Howland, DFO).

G) Great Bear Lake – *q-value* vs. depth habitat (fish from shallow and deep habitats are classified as Lake Cisco and Shortjaw by Howland et al. 2013)



H) Great Bear Lake – *q-value* vs. Gill Rakers in Dease Arm

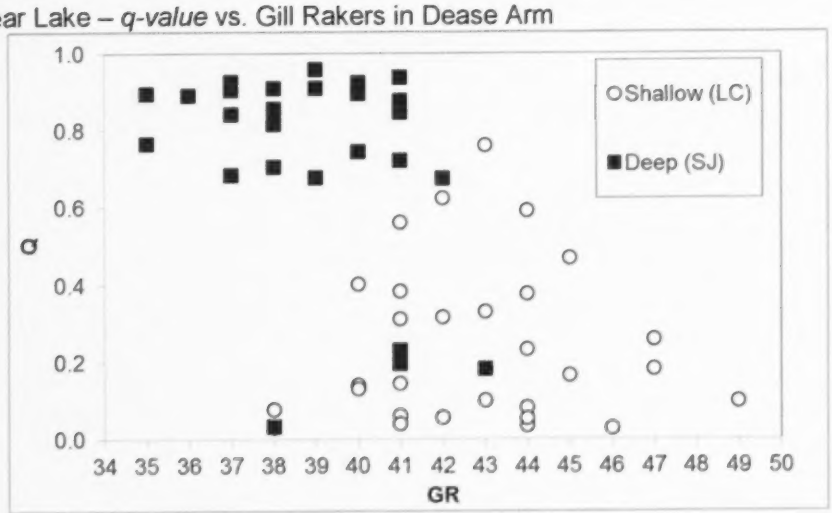


Figure 4(G-H). Association between individual genetic characteristics (q-value determined by Structure, expressed as the coefficient of ancestry of a genotype in the cluster best associated with SJ) and phenotypic attributes (DF1: Discriminant Function 1 score, GR: Total number of gill rakers, TL: Total body length in mm). Phenotypic attributes were from Murray (2006) or provided by researchers (S. Reid, OMNR; T. Pratt, DFO; K. Howland, DFO).

I) Great Bear Lake – *q-value* vs. Gill Rakers in Keith Arm

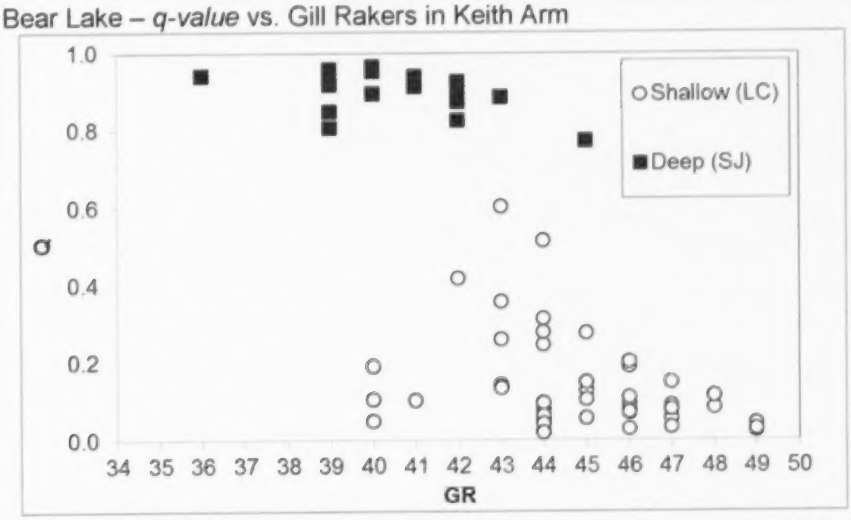


Figure 4(I). Association between individual genetic characteristics (q-value determined by Structure, expressed as the coefficient of ancestry of a genotype in the cluster best associated with SJ) and phenotypic attributes (DF1: Discriminant Function 1 score, GR: Total number of gill rakers, TL: Total body length in mm). Phenotypic attributes were from Murray (2006) or provided by researchers (S. Reid, OMNR; T. Pratt, DFO; K. Howland, DFO).